Deterministic, quenched, and annealed parameter estimation for heterogeneous network models

Marzio Di Vece,[1,2,*] Diego Garlaschelli,[1,3,4] and Tiziano Squartini[1,2,4,5]

[1]*IMT School for Advanced Studies, Piazza San Francesco 19, 55100 Lucca, Italy*
[2]*Scuola Normale Superiore, P.zza dei Cavalieri 7, 56126 Pisa, Italy*
[3]*Lorentz Institute for Theoretical Physics, University of Leiden, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands*
[4]*INdAM-GNAMPA Istituto Nazionale di Alta Matematica, 00185 Rome, Italy*
[5]*Institute for Advanced Study, University of Amsterdam, Oude Turfmarkt 145, 1012 GC Amsterdam, The Netherlands*

(Received 13 March 2023; revised 29 July 2023; accepted 20 September 2023; published 1 November 2023)

At least two different approaches to define and solve statistical models for the analysis of economic systems exist: the typical, econometric one, interpreting the gravity model specification as the expected link weight of an arbitrary probability distribution, and the one rooted in statistical physics, constructing maximum-entropy distributions constrained to satisfy certain network properties. In a couple of recent companion papers, they have been successfully integrated within the framework induced by the constrained minimization of the Kullback-Leibler divergence: specifically, two broad classes of models have been devised, i.e., the integrated and conditional ones, defined by different, probabilistic rules to place links, load them with weights and turn them into proper, econometric prescriptions. Still, the recipes adopted by the two approaches to estimate the parameters entering into the definition of each model differ. In econometrics, a likelihood that decouples the binary and weighted parts of a model, treating a network as deterministic, is typically maximized; to restore its random character, two alternatives exist: either solving the likelihood maximization on each configuration of the ensemble and taking the average of the parameters afterwards or taking the average of the likelihood function and maximizing the latter one. The difference between these approaches lies in the order in which the operations of averaging and maximization are taken—a difference that is reminiscent of the quenched and annealed ways of averaging out the disorder in spin glasses. The results of the present contribution, devoted to comparing these recipes in the case of continuous, conditional network models, indicate that the annealed estimation recipe represents the best alternative to the deterministic one.

DOI: 10.1103/PhysRevE.108.054301

I. INTRODUCTION

Over the last 20 years, the growth of network science has impacted several disciplines by establishing new, empirical facts about the structural properties of the related systems. Prominent examples are provided by economics and finance: the growing availability of data has motivated researchers to explore and model the architecture of cryptocurrencies [1], interbank networks [2], production networks [3], and trading networks [4–7].

Modeling the establishment of a connection and the corresponding weight simultaneously poses a serious challenge. Econometrics prescribes to estimate binary and weighted parameters either separately, within the context of hurdle models [8], or jointly, within the context of zero-inflated models [9]; in both cases, the gravity model specification [10] $\langle w_{ij}\rangle_{\text{GM}} = f(\omega_i, \omega_j, d_{ij}|\underline{\phi}) = e^{\rho}(\omega_i\omega_j)^{\alpha} d_{ij}^{\gamma}$—where $\omega_i \equiv \text{GDP}_i/\overline{\text{GDP}}$ is the GDP of country i divided by the arithmetic mean of the GDPs of all countries, d_{ij} is the geographic distance between the capitals of countries i and j, and $\underline{\phi} \equiv (\rho, \alpha, \gamma)$ is the vector of parameters defining the gravity model specification—is interpreted as the expected value of a probability distribution whose functional form is arbitrary. On the other hand, the approach rooted in statistical physics constructs maximum-entropy distributions, constrained to satisfy certain network properties [11–15].

In a couple of recent, companion papers [16,17] the two, aforementioned approaches have been integrated within the framework induced by the constrained optimization of the Kullback-Leibler (KL) divergence [18]. In particular, two broad classes of models have been constructed, i.e., the integrated and conditional ones, defined by different, probabilistic rules to place links, load them with weights, and turn them into properly econometric prescriptions. For what concerns integrated models, the first two rules follow from a single, constrained optimization of the KL divergence [19]; for what concerns conditional models, the two rules are disentangled and the functional form of the weight distribution follows from a conditional, optimization procedure [20]. Still, the prescriptions adopted by the two approaches to carry out the estimation of the parameters entering into the definition of each model differ.

*marzio.divece@imtlucca.it

The present contribution is devoted to comparing these recipes in the case of continuous, conditional network models defined by both homogeneous and heterogeneous constraints.

II. MINIMIZATION OF THE KULLBACK -LEIBLER DIVERGENCE

The functional form of continuous, conditional network models can be identified through the constrained minimization of the KL divergence of a distribution Q from a prior distribution R, i.e.,

$$D_{\mathrm{KL}}(Q||R)=\int_{\mathbb{W}}Q(\mathbf{W})\ln\frac{Q(\mathbf{W})}{R(\mathbf{W})}d\mathbf{W},\qquad(1)$$

where $\mathbf{W}$ is one of the possible values of a continuous random variable, $\mathbb{W}$ is the set of possible values that $\mathbf{W}$ can take, $Q(\mathbf{W})$ is the (multivariate) probability density function to be estimated, and $R(\mathbf{W})$ plays the role of prior distribution, whose divergence from $Q(\mathbf{W})$ must be minimized: in our setting, $\mathbf{W}$ represents an entire network whose weights now obey the property $w_{ij}\in\mathbb{R}_0^+$, $\forall\, i<j$. Such an optimization scheme embodies the so-called minimum discrimination information principle [16,17], implementing the idea that, as new information becomes available, an updated distribution $Q(\mathbf{W})$ should be chosen to make its discrimination from the prior distribution $R(\mathbf{W})$ as hard as possible.

Let us now separate both the prior and posterior distributions into a purely binary part and a conditional, weighted one; the positions $Q(\mathbf{W})=P(\mathbf{A})Q(\mathbf{W}|\mathbf{A})$ and $R(\mathbf{W})=T(\mathbf{A})R(\mathbf{W}|\mathbf{A})$, where $\mathbf{A}$ denotes the binary projection of the weighted network $\mathbf{W}$ (i.e., $\Theta[\mathbf{W}]=\mathbf{A}$), $T(\mathbf{A})$ represents the binary prior, and $R(\mathbf{W}|\mathbf{A})$ represents the conditional, weighted prior, lead the KL divergence to be rewritable as

$$D_{\mathrm{KL}}(Q||R)=D_{\mathrm{KL}}(P||T)+D_{\mathrm{KL}}(\overline{Q}||\overline{R}),\qquad(2)$$

i.e., as a sum of the two addenda

$$D_{\mathrm{KL}}(P||T)=\sum_{\mathbf{A}\in\mathbb{A}}P(\mathbf{A})\ln\frac{P(\mathbf{A})}{T(\mathbf{A})},\qquad(3)$$

$$D_{\mathrm{KL}}(\overline{Q}||\overline{R})=\sum_{\mathbf{A}\in\mathbb{A}}P(\mathbf{A})\int_{\mathbb{W}_{\mathbf{A}}}Q(\mathbf{W}|\mathbf{A})\ln\frac{Q(\mathbf{W}|\mathbf{A})}{R(\mathbf{W}|\mathbf{A})}d\mathbf{W}.\qquad(4)$$

In what follows, we will deal with completely uninformative priors, a choice that amounts at considering the (somehow, simplified) expression

$$-S(Q)=-S(P)-S(\overline{Q}|P),\qquad(5)$$

i.e., minus the joint entropy, where

$$S(P)=-\sum_{\mathbf{A}\in\mathbb{A}}P(\mathbf{A})\ln P(\mathbf{A})\qquad(6)$$

is the Shannon entropy of the probability distribution describing the binary projection of the network structure [14,15] and

$$S(\overline{Q}|P)=-\sum_{\mathbf{A}\in\mathbb{A}}P(\mathbf{A})\int_{\mathbb{W}_{\mathbf{A}}}Q(\mathbf{W}|\mathbf{A})\ln Q(\mathbf{W}|\mathbf{A})d\mathbf{W}\qquad(7)$$

is the conditional Shannon entropy of the probability distribution describing the weighted network structure [16,17,20]. Notice that, when continuous models are considered, $S(\overline{Q}|P)$ is defined by a sum running over all the binary configurations within the ensemble $\mathbb{A}$ and an integral over all the weighted configurations that are compatible with each, specific, binary structure, i.e., $\mathbb{W}_{\mathbf{A}}=\{\mathbf{W}:\Theta[\mathbf{W}]=\mathbf{A}\}$. For a more detailed discussion, see Appendix A.

The functional form of $P(\mathbf{A})$ can be determined by carrying out the usual, constrained maximization of Shannon entropy [14,15]; remarkably, any set of (binary) constraints considered in the present paper will lead to the same expression for $P(\mathbf{A})$, i.e., $P(\mathbf{A})=\prod_{i<j}p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}}$, with $p_{ij}=x_{ij}/(1+x_{ij})$: specifically, the position $x_{ij}\equiv x$ individuates the undirected binary random graph model (UBRGM), the position $x_{ij}\equiv x_ix_j$ individuates the undirected binary configuration model (UBCM), and the position $x_{ij}\equiv\delta\omega_i\omega_j$ individuates the logit model (LM) [21].

On the other hand, the functional form of $Q(\mathbf{W}|\mathbf{A})$ can be determined by carrying out the constrained maximization of $S(\overline{Q}|P)$, the set of constraints being, now,

$$1=\int_{\mathbb{W}_{\mathbf{A}}}P(\mathbf{W}|\mathbf{A})d\mathbf{W},\ \forall\,\mathbf{A}\in\mathbb{A},\qquad(8)$$

$$\langle C_\alpha\rangle=\sum_{\mathbf{A}\in\mathbb{A}}P(\mathbf{A})\int_{\mathbb{W}_{\mathbf{A}}}Q(\mathbf{W}|\mathbf{A})C_\alpha(\mathbf{W})d\mathbf{W},\ \forall\,\alpha;\qquad(9)$$

while the first condition ensures the normalization of the probability distribution, the vector $\{C_\alpha(\mathbf{W})\}$ represents the proper set of weighted constraints. The distribution induced by such an optimization problem reads

$$Q(\mathbf{W}|\mathbf{A})=\frac{e^{-H(\mathbf{W})}}{Z_{\mathbf{A}}}=\frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}_{\mathbf{A}}}e^{-H(\mathbf{W})}d\mathbf{W}}\qquad(10)$$

if $\mathbf{W}\in\mathbb{W}_{\mathbf{A}}$ and 0 otherwise. While the Hamiltonian $H(\mathbf{W})=\sum_\alpha\psi_\alpha C_\alpha(\mathbf{W})$ lists the constraints, the quantity at the denominator is the partition function, conditional on the fixed topology $\mathbf{A}$ [20].

For mathematical convenience, in what follows we will consider separable Hamiltonians, i.e., functions that can be written as sums of node pair-specific Hamiltonians: $H(\mathbf{W})=\sum_{i<j}H_{ij}(w_{ij})$; this choice leads to the result

$$Q(\mathbf{W}|\mathbf{A})=\frac{e^{-\sum_{i<j}H_{ij}(w_{ij})}}{\int_{\mathbb{W}_{\mathbf{A}}}e^{-\sum_{i<j}H_{ij}(w_{ij})}d\mathbf{W}}=\prod_{i<j}\frac{e^{-H_{ij}(w_{ij})}}{\left[\int_{m_{ij}}^{+\infty}e^{-H_{ij}(w_{ij})}dw_{ij}\right]^{a_{ij}}}=\prod_{i<j}\frac{e^{-H_{ij}(w_{ij})}}{\zeta_{ij}^{a_{ij}}}\qquad(11)$$

(with m_{ij} being the pair-specific, minimum weight allowed by a given model and ζ_{ij} being the corresponding partition function), irrespectively from the specific, functional form of $H_{ij}(w_{ij})$ [17]. For a more detailed discussion, see Appendix B.

III. ESTIMATION OF THE PARAMETERS

Several, alternative recipes are viable to estimate the parameters entering into the definition of continuous, conditional network models.

A. Deterministic parameter estimation

The simplest one prescribes to consider the traditional likelihood function

$$\begin{aligned}\ln Q(\mathbf{W}^*) &= \ln[P(\mathbf{A}^*)Q(\mathbf{W}^*|\mathbf{A}^*)]\\ &= \ln P(\mathbf{A}^*) + \ln Q(\mathbf{W}^*|\mathbf{A}^*),\end{aligned} \tag{12}$$

with $\mathbf{W}^*$ ($\mathbf{A}^*$) being the empirical, weighted (binary) adjacency matrix; its maximization allows the parameters entering into the definition of the purely topological distribution and those entering into the definition of the conditional weighted one to be estimated in a totally disentangled fashion [17], in fact, maximizing

$$\begin{aligned}\mathcal{L}_{\underline{\psi}} &= \ln Q(\mathbf{W}^*|\mathbf{A}^*)\\ &= -H(\mathbf{W}^*) - \ln Z_{\mathbf{A}^*}\\ &= -H(\mathbf{W}^*) - \ln\left[\int_{\mathbb{W}_{\mathbf{A}^*}} e^{-H(\mathbf{W})}d\mathbf{W}\right],\end{aligned} \tag{13}$$

with respect to the unknown parameters leads us to find the vector of values $\underline{\psi}^*$ satisfying the vector of relationships

$$\langle\mathbf{C}\rangle_{\mathbf{A}^*}(\underline{\psi}^*) \equiv \mathbf{C}^*, \tag{14}$$

which stands for the set of relationships $\langle C_\alpha\rangle_{\mathbf{A}^*}(\underline{\psi}^*) \equiv C^*_\alpha$, $\forall\,\alpha$, each one equating the model-induced average value of the corresponding constraint to its empirical value, marked with an asterisk.

This first approach to parameter estimation can be named deterministic to stress that $\mathbf{A}^*$ is considered as not being subject to variation; otherwise stated, this recipe—which is the most common in econometrics—prescribes estimating the parameters entering into the definition of the conditional, weighted probability distribution by assuming the network topology to be fixed.

B. Annealed parameter estimation

Topology, however, is a random variable itself, obeying the probability distribution $P(\mathbf{A})$. As a consequence, the deterministic recipe for parameter estimation could lead to inconsistencies, should the description of $\mathbf{A}^*$ provided by $P(\mathbf{A})$ not be accurate. The variability induced by $P(\mathbf{A})$ can be properly accounted for by considering the generalized likelihood [20]

$$\begin{aligned}\mathcal{G}_{\underline{\psi}} &= \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\ln Q(\mathbf{W}^*|\mathbf{A})\\ &= \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})[-H(\mathbf{W}^*) - \ln Z_{\mathbf{A}}]\\ &= -H(\mathbf{W}^*) - \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\ln\left[\int_{\mathbb{W}_{\mathbf{A}^*}} e^{-H(\mathbf{W})}d\mathbf{W}\right] = \langle\mathcal{L}_{\underline{\psi}}\rangle,\end{aligned} \tag{15}$$

whose maximization leads us to find the vector of values $\underline{\psi}^*$ satisfying the vector of relationships

$$\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\langle\mathbf{C}\rangle_{\mathbf{A}}(\underline{\psi}^*) = \langle\mathbf{C}\rangle(\underline{\psi}^*) = \mathbf{C}^*, \tag{16}$$

which stands for the set of relationships $\langle C_\alpha\rangle(\underline{\psi}^*) \equiv C^*_\alpha$, $\forall\,\alpha$. Taking this average is conceptually similar to taking the annealed average in physics: parameter estimation is carried out while random variables—again, the entries of the adjacency matrix—are left to vary.

Interestingly, the deterministic recipe is a special case of the annealed recipe since the former can be recovered by posing $P(\mathbf{A}) \equiv \delta_{\mathbf{A},\mathbf{A}^*}$: in this case, in fact,

$$\begin{aligned}\mathcal{G}_{\underline{\psi}} &= -H(\mathbf{W}^*) - \sum_{\mathbf{A}\in\mathbb{A}} \delta_{\mathbf{A},\mathbf{A}^*}\ln Z_{\mathbf{A}}\\ &= -H(\mathbf{W}^*) - \ln Z_{\mathbf{A}^*} = \mathcal{L}_{\underline{\psi}};\end{aligned} \tag{17}$$

similarly, $\sum_{\mathbf{A}\in\mathbb{A}}\delta_{\mathbf{A},\mathbf{A}^*}\langle\mathbf{C}\rangle_{\mathbf{A}}(\underline{\psi}^*) = \langle\mathbf{C}\rangle_{\mathbf{A}^*}(\underline{\psi}^*) = \mathbf{C}^*$.

C. Quenched parameter estimation

A viable alternative to properly account for the variability induced by $P(\mathbf{A})$ is that of reversing the two operations of likelihood maximization and ensemble averaging: in other words, one can (1) numerically sample the ensemble of configurations induced by $P(\mathbf{A})$, (2) maximize the likelihood $\ln Q(\mathbf{W}^*|\mathbf{A})$ for each, generated network, and (3) take the average of the resulting set of parameters, according to the formula

$$\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\underline{\psi}^*(\mathbf{A}) = \langle\underline{\psi}^*\rangle, \tag{18}$$

the estimation of the αth parameter being assumed to coincide with the average $\langle\psi^*_\alpha\rangle$.

Taking this average is conceptually similar to taking the quenched average in physics: random variables—in the specific case, the entries of the adjacency matrix—are frozen, parameter estimation is carried out and, only at the end, the values of the parameters are averaged over the ensemble of configurations induced by $P(\mathbf{A})$.

As our models inherit their functional form from the constrained minimization of the KL divergence, each parameter controls for a specific constraint: When employing the deterministic recipe, such a circumstance makes each parameter configuration dependent; when employing either the annealed or the quenched recipe, instead, accounting for the variability of a network structure induces a sort of loss of memory about its empirical, purely topological details.

IV. RESULTS

To test if the deterministic, annealed, and quenched prescriptions lead to the same estimation, let us focus on a number of variants of the conditional exponential model (CEM), induced by the positions $H^{\text{CEM}}_{ij} = \beta_{ij}w_{ij}$ and $\zeta^{\text{CEM}}_{ij} = \beta^{-1}_{ij}$:

$$\begin{aligned}Q(\mathbf{W}) &= P(\mathbf{A})Q(\mathbf{W}|\mathbf{A})\\ &= \prod_{i<j} p^{a_{ij}}_{ij}(1-p_{ij})^{1-a_{ij}}\prod_{i<j}\beta^{a_{ij}}_{ij}e^{-\beta_{ij}w_{ij}};\end{aligned} \tag{19}$$

naturally, $q_{ij}(w_{ij}=0|a_{ij}=0) = 1$ (i.e., if nodes i and j are not connected, the weight of the corresponding link is zero with probability equal to one) and $q_{ij}(w_{ij}>0|a_{ij}=1) = \beta_{ij}e^{-\beta_{ij}w_{ij}}$.

In what follows, we will consider three different instances of $p_{ij} = x_{ij}/(1 + x_{ij})$, corresponding to

(1) The UBRGM, defined by posing $x_{ij} \equiv x$ and induced by the maximization of $S(P)$ while constraining the total number of links, $L(\mathbf{A}^*) \equiv L^* = \sum_{i<j} a^*_{ij}$, i.e.,

$$p_{ij}^{\text{UBRGM}} \equiv \frac{x}{1+x}. \tag{20}$$

(2) The UBCM, defined by posing $x_{ij} \equiv x_i x_j$ and induced by the maximization of $S(P)$ while constraining the whole degree sequence, $\{k_i(\mathbf{A}^*)\}_{i=1}^N \equiv \{k_i^*\}_{i=1}^N$, with $k_i^* = \sum_{j(\neq i)} a^*_{ij}$, i.e.,

$$p_{ij}^{\text{UBCM}} \equiv \frac{x_i x_j}{1 + x_i x_j}. \tag{21}$$

(3) Two different instances of the LM, both representing a fitness-driven version of the UBCM, (again) induced by constraining the total number of links, $L(\mathbf{A}^*) \equiv L^* = \sum_{i<j} a^*_{ij}$. The first one is defined by posing $x_{ij} \equiv \delta \omega_i \omega_j$, i.e.,

$$p_{ij}^{\text{LM}} \equiv \frac{\delta \omega_i \omega_j}{1 + \delta \omega_i \omega_j}, \tag{22}$$

and has been employed to study the year 2017 of the CEPII-BACI version of the World Trade Web (WTW) [22], that is, a network of $N = 171$ nodes and a link density of $d = 0.87$. The second one is defined by posing $x_{ij} \equiv \delta s_i s_j$, i.e.,

$$p_{ij}^{\text{LM}} = \frac{\delta s_i s_j}{1 + \delta s_i s_j}, \tag{23}$$

and has been employed to study a snapshot of the Bitcoin Lightning Network (BLN) taken on 01/03/2019 [23], that is, a network of $N = 5012$ nodes and a link density of $d = 0.003$.

A. Scalar variant of the conditional exponential model

Let us start by considering the scalar or homogeneous variant of the CEM, defined by the position $\beta_{ij} \equiv \beta, \forall\, i < j$.

In this case, the deterministic recipe for parameter estimation prescribes to maximize the likelihood

$$\mathcal{L}_{\underline{\psi}} = \sum_{i<j} [-\beta w^*_{ij} + a^*_{ij} \ln \beta] = -\beta W^* + L^* \ln \beta, \tag{24}$$

where $W(\mathbf{W}^*) \equiv W^* = \sum_{i<j} w^*_{ij}$ and whose optimization leads to the expression $\beta = L^*/W^*$. The annealed recipe prescribes to maximize the likelihood

$$\mathcal{G}_{\underline{\psi}} = \sum_{i<j} [-\beta w^*_{ij} + p_{ij} \ln \beta] = -\beta W^* + \langle L \rangle \ln \beta, \tag{25}$$

whose optimization leads to the expression $\beta = \langle L \rangle / W^*$. The quenched recipe, on the other hand, prescribes to calculate the average

$$\langle \beta \rangle = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \beta(\mathbf{A}) = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \frac{L(\mathbf{A})}{W^*} = \frac{\langle L \rangle}{W^*}, \tag{26}$$

since, now, $\beta(\mathbf{A}) = L(\mathbf{A})/W^*$.

In the case of the scalar variant of the CEM, the estimations coincide for any null model preserving the total number of links, i.e., ensuring that $\langle L \rangle = L^*$, regardless of the network density. Such a result is confirmed by Fig. 1, where each recipe has been implemented on the WTW, by adopting the distributions induced by the UBRGM (blue), the UBCM (green), and the LM (red). Specifically, the deterministic estimation (black, solid line) and the annealed estimations (blue, green, and red solid lines) overlap; moreover, each annealed estimation overlaps with the the corresponding, quenched estimation, i.e., the average value of the related, quenched distribution (blue, green, and red dash-dotted lines).

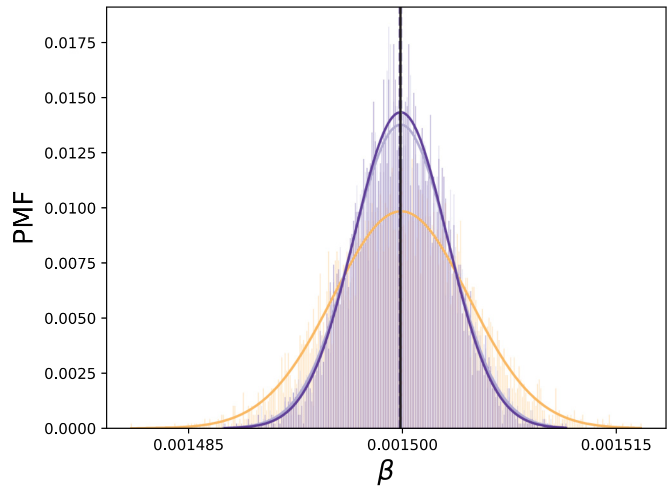

FIG. 1. Estimations of the parameter β, entering the definition of the homogeneous version of the CEM, where the binary topology is either deterministic (black vertical line) or generated via the UBRGM (light orange or light grey), the UBCM (purple or dark grey), and the LM (light purple or grey). The deterministic approach leads to a single estimate, while the other approaches lead to either a single annealed estimate (vertical, solid lines) or to a whole distribution of quenched estimates (empirical distribution constructed over an ensemble of 5.000 binary configurations with theoretical curves, binomial or Poisson-binomial, dependent on the binary model; the corresponding average value is indicated by a vertical, dash-dotted line). The annealed parameter estimates, the average values of the quenched parameter distributions and the deterministic parameter estimate coincide. Data refers to the year 2017 of the CEPII-BACI version of the WTW [22].

In the case of the UBRGM-induced homogeneous version of the CEM, the quenched distribution of the parameter $\beta(\mathbf{A}) = L(\mathbf{A})/W^*$ inherits the distribution of the total number of links, i.e., $L \sim \text{Bin}(N(N-1)/2, p)$, with $p = 2L^*/N(N-1)$: more precisely, $W\beta \sim \text{Bin}(N(N-1)/2, p)$; analogously for the UBCM- and the LM-induced homogeneous versions of the CEM—the only difference being that, now, L obeys two, different, Poisson-binomial (PB) distributions.

B. Vector variant of the conditional exponential model

Let us now consider the vector or weakly heterogeneous variant of the CEM, defined by the position $\beta_{ij} \equiv \beta_i + \beta_j$, $\forall\, i < j$.

In this case, the deterministic recipe for parameter estimation prescribes to maximize the likelihood

$$\begin{aligned}\mathcal{L}_{\underline{\psi}} &= \sum_{i<j} [-(\beta_i + \beta_j) w^*_{ij} + a^*_{ij} \ln(\beta_i + \beta_j)] \\ &= -\sum_i \beta_i s^*_i + \sum_{i<j} a^*_{ij} \ln(\beta_i + \beta_j),\end{aligned} \tag{27}$$

where $s_i(\mathbf{W}^*) \equiv s_i^* = \sum_{j(\neq i)} w_{ij}^*$ and whose optimization requires us to solve the system of equations

$$s_i^* = \sum_{j(\neq i)} \frac{a_{ij}^*}{\beta_i + \beta_j}, \ \forall\, i. \tag{28}$$

The annealed recipe, instead, prescribes to maximize the likelihood

$$\begin{aligned}\mathcal{G}_{\underline{\psi}} &= \sum_{i<j} [-(\beta_i + \beta_j) w_{ij}^* + p_{ij} \ln(\beta_i + \beta_j)] \\ &= -\sum_i \beta_i s_i^* + \sum_{i<j} p_{ij} \ln(\beta_i + \beta_j),\end{aligned} \tag{29}$$

whose optimization requires us to solve the system of equations

$$s_i^* = \sum_{j(\neq i)} \frac{p_{ij}}{\beta_i + \beta_j}, \ \forall\, i \tag{30}$$

(notice that both the deterministic and the annealed version of the vector variant of the CEM are alternative instances of the so-called CReM$_\text{A}$, introduced in Ref. [20]). The quenched recipe, on the other hand, requires us to solve the system of equations $\langle \beta_i \rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \beta_i(\mathbf{A})$, $\forall\, i$ which no longer have an explicit expression. Devising some sort of approximation is, however, possible. Let us start by rewriting Eq. (30) as

$$\beta_i = \frac{1}{s_i^*} \sum_{j(\neq i)} \frac{p_{ij}}{1 + \beta_j/\beta_i}, \ \forall\, i \tag{31}$$

and consider the node whose coefficient is the largest one. This allows us to write $\beta_i \simeq \sum_{j(\neq i)} p_{ij}/s_i^* = \langle k_i \rangle / s_i^*$: in the case we implemented the UBRGM, we would obtain $\beta_i(\mathbf{A}) \simeq 2L(\mathbf{A})/N s_i^*$, hence expecting the quenched distribution of $N s_i^* \beta_i / 2$ to coincide with $\text{Bin}(N(N-1)/2, p)$; if, on the other hand, we implemented the UBCM, we would obtain $\beta_i(\mathbf{A}) \propto k_i(\mathbf{A})/s_i^*$, hence expecting the quenched distribution of $s_i^* \beta_i$ to obey a PB. Again, the estimations coincide for any null model preserving the structural properties characterising the binary recipe implemented.

More generally, the mutual relationships between the estimations provided by the three recipes are node dependent (see Fig. 2, illustrating the case study of node 166 of the WTW and Fig. 4 in Appendix C): In general, however, each annealed estimation overlaps with the average value of the related quenched distribution. Moreover, the deterministic estimation is very close to the UBCM-induced, annealed one; such a result is a consequence of the accurate description of the empirical network topology provided by the UBCM—in fact, much more accurate than the ones provided by the UBRGM and the LM: indeed, the better the approximation $p_{ij} \simeq a_{ij}$, $\forall\, i < j$, the closer the annealed estimation to the deterministic one.

This is even more evident when considering the tensor variant of the CEM, in which case the three optimization procedures lead to the expressions $\beta_\text{det} = a_{ij}^*/\hat{w}_{ij}, \forall\, i < j$, and $\beta_\text{ann} = \langle \beta \rangle_\text{que} = p_{ij}/\hat{w}_{ij}, \forall\, i < j$, with $\hat{w}_{ij}$ representing an estimate of the empirical weight w_{ij}^*; if, however, $\hat{w}_{ij} \equiv w_{ij}^*$, $\forall\, i < j$ then, for consistency, $p_{ij} \equiv a_{ij}^*$ and the three recipes coincide.

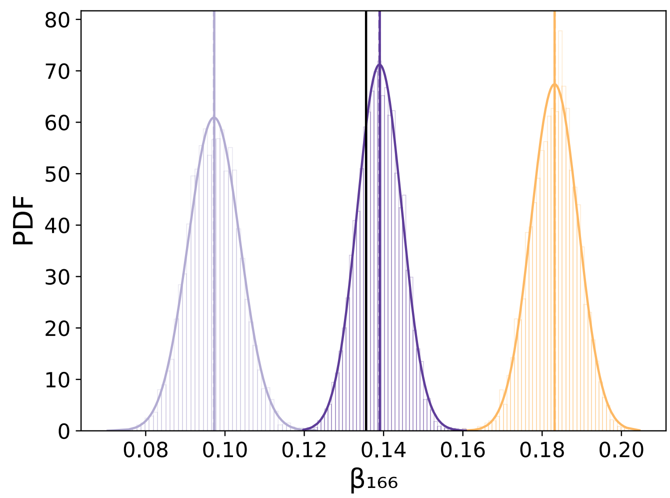

FIG. 2. Estimations of the parameter β_{166} entering the definition of the weakly heterogeneous version of the CEM, where the binary topology is either deterministic (black vertical line) or generated via the UBRGM (light orange or light grey), the UBCM (purple or dark grey), and the LM (light purple or grey). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, annealed estimate (vertical, solid lines) or to a whole distribution of quenched estimates (histograms with normal density curves having the same average and standard deviation, constructed over an ensemble of 5.000 binary configurations; the average value is indicated by a vertical dash-dotted line). Each annealed parameter estimate coincides with the average value of the corresponding quenched distribution although the distributions induced by the three binary recipes are well separated. In addition, the deterministic parameter estimate is very close to the UBCM-induced, annealed one. Data refers to the year 2017 of the CEPII-BACI version of the WTW [22].

C. Econometric variant of the conditional exponential model

As a third case study, let us focus on the econometric variant of the CEM, defined by posing $\beta_{ij} \equiv \beta_0 + z_{ij}^{-1}$, $\forall\, i < j$, where $z_{ij} \equiv e^\rho (\omega_i \omega_j)^\alpha d_{ij}^\gamma$ represents the gravity model specification traditionally employed to analyze undirected, weighted trade networks and β_0 is a structural parameter to be tuned to ensure that $\langle W \rangle = W^*$. In this case, the deterministic recipe for parameter estimation prescribes to maximize the likelihood

$$\mathcal{L}_{\underline{\psi}} = \sum_{i<j} \left[-\left(\beta_0 + z_{ij}^{-1}\right) w_{ij}^* + a_{ij}^* \ln\left(\beta_0 + z_{ij}^{-1}\right)\right], \tag{32}$$

whose optimization requires us to solve the system of equations

$$W^* = \sum_{i<j} \frac{a_{ij}^*}{\beta_0 + z_{ij}^{-1}}, \tag{33}$$

$$\sum_{i<j} w_{ij}^* \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}} = \sum_{i<j} \frac{a_{ij}^*}{\beta_0 + z_{ij}^{-1}} \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}}. \tag{34}$$

The annealed recipe, instead, prescribes to maximize the likelihood

$$\mathcal{G}_{\underline{\psi}} = \sum_{i<j} \left[-\left(\beta_0 + z_{ij}^{-1}\right) w_{ij}^* + p_{ij} \ln\left(\beta_0 + z_{ij}^{-1}\right)\right], \tag{35}$$

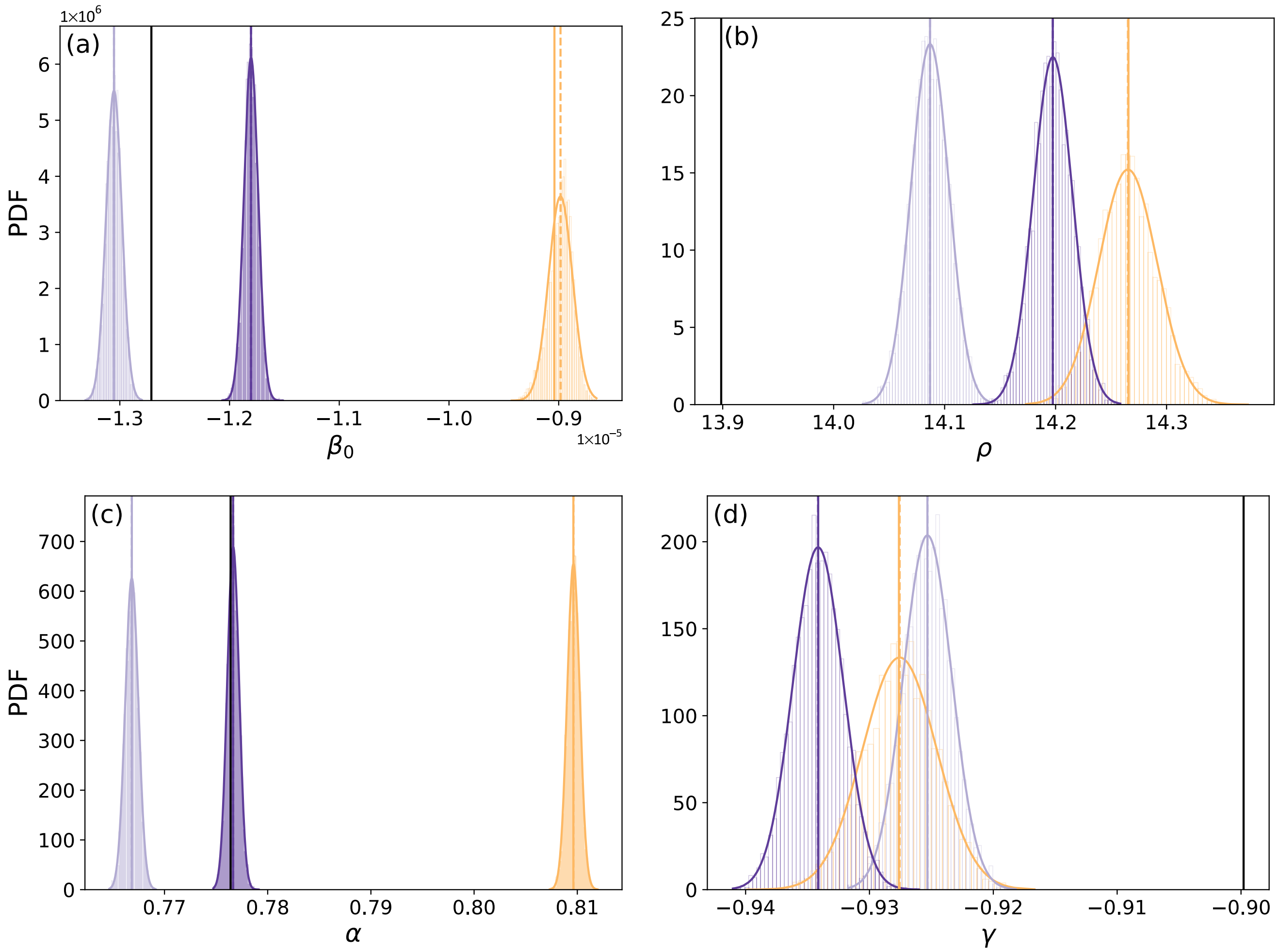

FIG. 3. Estimations of the parameters (a) β_0, (b) ρ, (c) α, and (d) γ, entering the definition of the econometric version of the CEM, where the binary topology is either deterministic (black vertical line) or generated via the UBRGM (light orange or light grey), the UBCM (purple or dark grey), and the LM (light purple or grey). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, annealed estimate (vertical, solid lines) or to a whole distribution of quenched estimates (histograms with kernel density curves, constructed over an ensemble of 5.000 binary configurations; the corresponding average value is indicated by a vertical, dash-dotted line). Each annealed parameter estimate coincides with the average value of the corresponding quenched distribution, although the distributions induced by the three binary recipes may overlap or not; the deterministic estimate, instead, overlaps with the other, two ones only for the parameter α, under the UBCM-induced, binary recipe. Data refers to the year 2017 of the CEPII-BACI version of the WTW [22].

whose optimization requires us to solve the system of equations

$$W^* = \sum_{i<j} \frac{p_{ij}}{\beta_0 + z_{ij}^{-1}}, \tag{36}$$

$$\sum_{i<j} w_{ij}^* \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}} = \sum_{i<j} \frac{p_{ij}}{\beta_0 + z_{ij}^{-1}} \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}}. \tag{37}$$

The quenched recipe, on the other hand, requires us to solve the system of equations $\langle \beta_0 \rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\beta_0(\mathbf{A})$ and $\langle \underline{\phi} \rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\underline{\phi}(\mathbf{A})$, which no longer have an explicit expression.

Figures 3 and 5 in Appendix C illustrate the case study of the WTW: although the quenched distributions induced by the three binary recipes are characterized by different shapes that may overlap (as in the case of the parameters ρ—under the UBRGM-induced and UBCM-induced binary recipes—and γ—under all, binary recipes) or not (as in the case of the parameters β_0 and α), annealed and quenched estimations always coincide (the only small discrepancy being observable for the parameter β_0, under the UBRGM-induced binary recipe). The deterministic estimation, instead, is compatible with the other two ones only for parameter α, under the UBCM-induced, binary recipe.

Sparse networks deserve a separate discussion. The results concerning the homogeneous and econometric variants of the BLN, defined by posing $\beta_{ij} \equiv \beta_0 + z_{ij}^{-1}$, $\forall\, i < j$, with $z_{ij} \equiv e^{\rho}(s_i s_j)^{\alpha}$, are analogous to the ones shown for the WTW—in the latter case, the annealed estimates of β_0, ρ, and α are very close to their quenched counterparts, the relative error $\text{RE} = |(\phi_i^{\text{ann}} - \phi_i^{\text{que}})/\phi_i^{\text{ann}}|$ amounting at $\simeq 10^{-3}$ for β_0 and $\simeq 10^{-4}$ for ρ, α. On the contrary, these conclusions no longer hold true when the weakly heterogeneous variant of the CEM is considered: in this case, in fact, carrying out the quenched approach can lead to binary configurations with disconnected nodes, a circumstance that impairs the correct estimation of

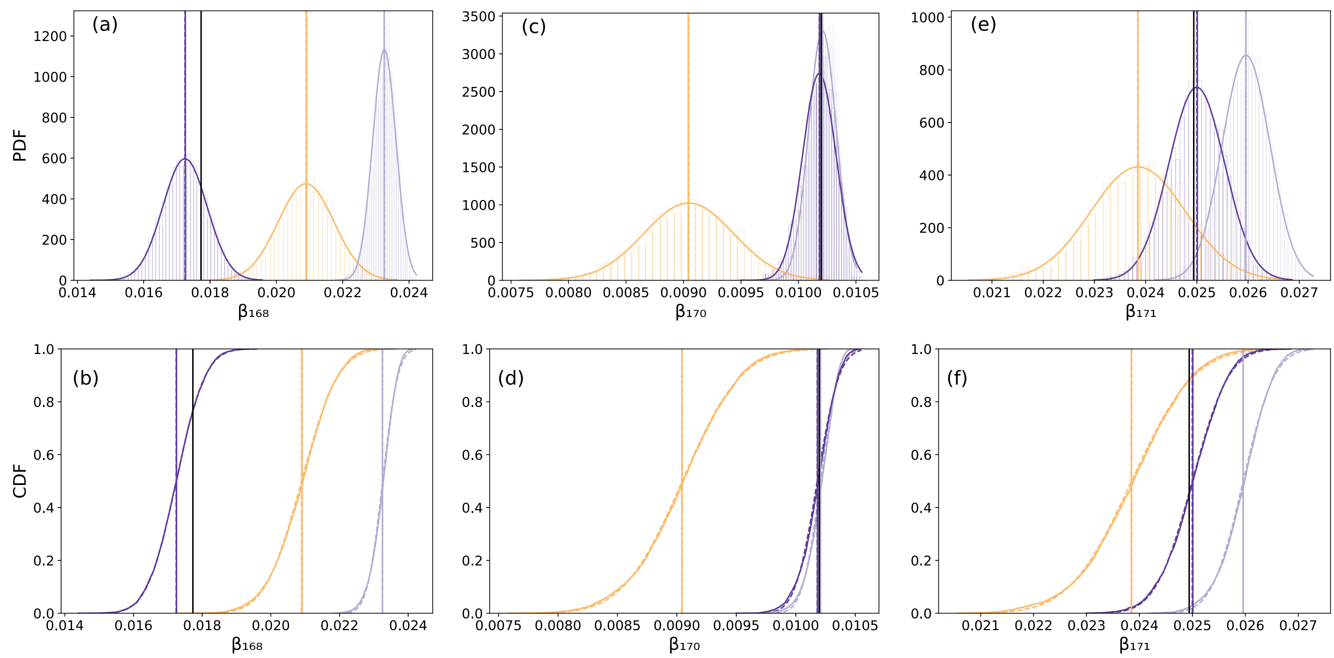

FIG. 4. Estimations of the parameters (a), (b) β_{168}; (c), (d) β_{170}; and (e), (f) β_{171}, entering the definition of the weakly heterogeneous version of the CEM, where the binary topology is either deterministic (black vertical line) or generated via the UBRGM (light orange or light grey), the UBCM (purple or dark grey), and the LM (light purple or grey). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, annealed estimate (vertical, solid lines) or to a whole distribution of quenched estimates (histograms with normal density curves having the same average and standard deviation, constructed over an ensemble of 5.000 binary configurations; the average value is indicated by a vertical, dash-dotted line). Each annealed estimate overlaps with the average value of the related quenched distribution, although (1) the latter ones are well separated in the case of node 168, (2) only partly overlapped in the case of node 171, and (3) the UBCM-induced and the LM-induced ones overlap while the UBRGM-induced one remains well separated in the case of node 170. Moreover, the deterministic estimates are always very close to (if not overlapping with) the UBCM-induced annealed ones. Although the empirical and theoretical CDFs (respectively depicted as solid lines and dotted lines in the bottom panels) seem to be in a very good agreement, the Anderson-Darling test never rejects the normality hypothesis only for node 166 and does not reject the normality hypothesis in the case of the UBCM-induced distribution of estimates for node 168.

the corresponding parameters; carrying out the annealed estimation, instead, remains a feasible task.

V. DISCUSSION

The present contribution focuses on three recipes for estimating the parameters entering into the definition of statistical network models, i.e., the deterministic, annealed, and quenched ones. To implement them, we have considered several variants of the CEM, i.e., the homogeneous one (defined by one, global parameter), the weakly heterogeneous one (defined by N, local parameters), and the econometric one (defined by four, global parameters), each one combined with three different recipes for estimating the network topology (i.e., the UBRGM, the UBCM, and the LM).

The deterministic recipe, routinely employed in econometrics to determine the so-called hurdle models [8], prescribes estimating the parameters associated to the weighted constraints on the empirical realisation of the network topology. Since it considers $\mathbf{A}^*$ as not being subject to variation, its use is recommended whenever $\text{Var}[a_{ij}] = p_{ij}(1 - p_{ij}) \simeq 0$ or, equivalently, $p_{ij} \simeq a_{ij}$, $\forall\, i < j$, i.e., whenever the binary random variables can be safely considered as deterministic or, more in general, whenever their (scale of) variation is negligible with respect to the (scale of) variation of the weighted random variables.

Accounting for such a variability in a fully consistent manner can be achieved upon adopting either the annealed recipe (according to which parameters are estimated on the average network topology) or the quenched recipe (according to which parameters are, first, estimated on a large number of binary configurations and, then, averaged); the main difference between these procedures lies in the order in which the two operations of averaging (of the entries of the binary adjacency matrix) and maximization (of the related likelihood function) are taken. Interestingly, no variant of the CEM is sensitive to this choice (neither the purely structural ones nor the econometric one); while, however, the coincidence of the annealed and quenched estimates for purely structural models can be explicitly verified, this is no longer true when the econometric variant is considered: in this case, in fact, one can proceed only numerically.

This evidence reveals the main limitation of the quenched approach, i.e., the need for resorting upon an explicit sampling of the chosen, binary ensemble. As any good sampling algorithm must lead to a faithful representation of the parent distribution, we are left with the following question: Is this always guaranteed, in all cases of interest to us?

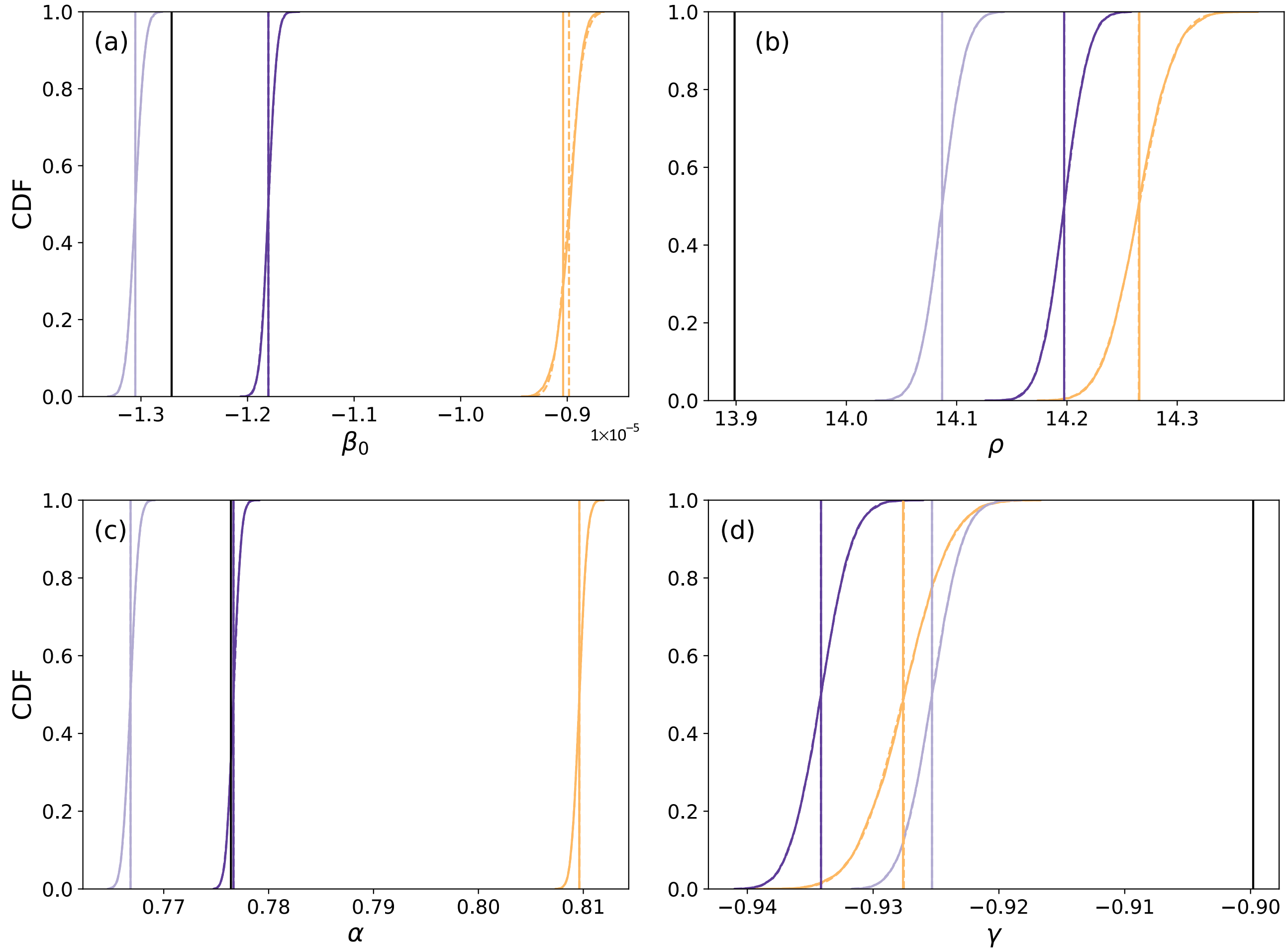

FIG. 5. Empirical CDFs for the parameters (a) β_0, (b) ρ, (c) α, and (d) γ entering the definition of the econometric version of the CEM, where the binary topology is either deterministic (black vertical line) or generated via the UBRGM (light orange or light grey), the UBCM (purple or dark grey), and the LM (light purple or grey). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, annealed estimate (vertical, solid lines) or to a whole distribution of quenched estimates (constructed over an ensemble of 5.000 binary configurations; the corresponding average value is indicated by a vertical, dash-dotted line). The shapes of the quenched, cumulative distributions induced by the three binary recipes are very similar.

This seems to be the case for dense networks. As shown in Ref. [24], a study of the coefficient of variation of the constraints defining the vector variant of the CEM (i.e., the ratio between standard deviation and the expected value of each degree) reveals it to vanish in the asymptotic limit: in other words, the fluctuations affecting each degree vanish, a result guaranteeing that the degree sequence of any configuration in the ensemble remains close enough to the empirical one.

When sparse networks are, instead, considered, the coefficient of variation of the constraints defining the vector variant of the CEM remains finite in the asymptotic limit: in other words, the fluctuations affecting each degree do not vanish, a result implying that the degree sequence of any configuration in the ensemble may largely differ from the empirical one; to provide a concrete example, nodes whose empirical degree is small may disconnect, hence inducing the resolution of a system of equations which is not even compatible with the set of constraints defining the original problem. Overcoming such a limitation implies quantifying the bias affecting the estimates in cases like these: although possible, calculations of this kind are far beyond the scope of the present paper.

Overall, then, two alternatives exist to overcome the main limitation of the deterministic estimation recipe, i.e., that of ignoring the variety of structures that are compatible with a given probability distribution $P(\mathbf{A})$, namely, the annealed and quenched ones. As the quenched recipe requires an explicit sampling of the ensemble—potentially leading to inconsistent estimates for sparse configurations—we believe the annealed one to represent the better alternative, (1) being *unbiased* by definition, (2) being *convenient* from a numerical point of view, and (3) reducing to the deterministic recipe in case the empirical configuration is not subject to variation.

ACKNOWLEDGMENTS

SoBigData.it receives funding from European Union – NextGenerationEU – National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) – Project: “SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics” – Prot. IR0000013

– Avviso n. 3264 del 28/12/2021. This work is also supported by PNRR-M4C2-Investimento 1.3, Partenariato Esteso PE00000013—FAIR-Future Artificial Intelligence Research—Spoke 1 Human-centered AI, funded by the European Commission under the NextGeneration EU program and by the project Network Analysis of Economic and Financial Resilience, Italian DM No. 289, 25-03-2021 (PRO3 Scuole) CUP D67G22000130001. D.G. acknowledges support from the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands) and the Netherlands Organization for Scientific Research (NWO/OCW). M.D.V. acknowledges support by the European Community programme under the funding schemes: ERC-2018-ADG G.A. 834756 "XAI: Science and technology for the eXplanation of AI decision making." M.D.V. and D.G. also acknowledge support from the Programma di Attività Integrata (PAI) project Prosociality, Cognition and Peer Effects (Pro.Co.P.E.), funded by IMT School for Advanced Studies Lucca.

APPENDIX A: CONDITIONAL NETWORK MODELS FROM KL DIVERGENCE MINIMIZATION

Discrete maximum-entropy models can be derived by performing a constrained maximization of Shannon entropy [11,12]. Here, however, we focus on continuous probability distributions: In such a case, mathematical problems are known to affect the definition of Shannon entropy as well as the resulting inference procedure. To restore the framework, one has to consider the KL divergence $D_{\text{KL}}(Q||R)$ of a distribution $Q(\mathbf{W})$ from a prior distribution $R(\mathbf{W})$ and reinterpret the maximization of the entropy associated to $Q(\mathbf{W})$ as the minimization of its distance from $R(\mathbf{W})$. Such an optimization scheme embodies the so-called minimum discrimination information principle, originally proposed by Kullback and Leibler [18] and requiring new data to produce an information gain that is as small as possible. In formulas, the KL divergence is defined as

$$D_{\text{KL}}(Q||R) = \int_{\mathbb{W}} Q(\mathbf{W}) \ln \frac{Q(\mathbf{W})}{R(\mathbf{W})} d\mathbf{W}; \tag{A1}$$

the class of conditional models can be introduced upon rewriting the posterior distribution $Q(\mathbf{W})$ as $Q(\mathbf{W}) = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A})$, where $\mathbf{A}$ denotes the binary projection of the weighted network $\mathbf{W}$. This equation allows us to split the KL divergence into the sum of three terms reading

$$D_{\text{KL}}(Q||R) = S(Q, R) - S(P) - S(\overline{Q}|P), \tag{A2}$$

where

$$S(Q, R) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln R(\mathbf{W}) d\mathbf{W} \tag{A3}$$

is the cross entropy, quantifying the amount of information required to identify a weighted network sampled from the distribution $Q(\mathbf{W})$ by employing the distribution $R(\mathbf{W})$,

$$S(P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln P(\mathbf{A}) \tag{A4}$$

is the Shannon entropy of the probability distribution describing the binary projection of the network structure, and

$$S(\overline{Q}|P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln Q(\mathbf{W}|\mathbf{A}) d\mathbf{W} \tag{A5}$$

is the conditional Shannon entropy of the probability distribution of the weighted network structure, given the binary projection. The expression for $S(Q, R)$ can be further manipulated as follows: Upon separating the prior distribution itself into a purely binary part and a conditional, weighted one, we can pose $R(\mathbf{W}) = T(\mathbf{A})R(\mathbf{W}|\mathbf{A})$, an expression that allows the KL divergence to be rewritten as

$$D_{\text{KL}}(Q||R) = +D_{\text{KL}}(P||T) + D_{\text{KL}}(\overline{Q}||\overline{R}), \tag{A6}$$

i.e., as a sum of the two addenda

$$D_{\text{KL}}(P||T) = +\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln \frac{P(\mathbf{A})}{T(\mathbf{A})}, \tag{A7}$$

$$D_{\text{KL}}(\overline{Q}||\overline{R}) = +\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln \frac{Q(\mathbf{W}|\mathbf{A})}{R(\mathbf{W}|\mathbf{A})} d\mathbf{W}, \tag{A8}$$

with $T(\mathbf{A})$ representing the binary prior and $R(\mathbf{W}|\mathbf{A})$ representing the conditional, weighted one. Dealing with completely uninformative priors amounts to considering the expression

$$-S(Q) = -S(P) - S(\overline{Q}|P), \tag{A9}$$

i.e., minus the joint entropy. The (independent) constrained optimization of $S(P)$ and $S(\overline{Q}|P)$ represents the starting point for deriving the members of the class of conditional models.

APPENDIX B: CONDITIONAL NETWORK MODELS: DETERMINING THE FUNCTIONAL FORM

The constrained maximization of $S(\overline{Q}|P)$ proceeds by specifying the set of weighted constraints reading

$$1 = \int_{\mathbb{W}_{\mathbf{A}}} P(\mathbf{W}|\mathbf{A}) d\mathbf{W}, \ \forall\, \mathbf{A} \in \mathbb{A}, \tag{B1}$$

$$\langle C_\alpha \rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}) d\mathbf{W}, \ \forall\, \alpha, \tag{B2}$$

the first condition ensuring the normalization of the probability distribution and the vector $\{C_\alpha(\mathbf{W})\}$ representing the proper set of weighted constraints. The distribution induced by them reads

$$\begin{aligned} Q(\mathbf{W}|\mathbf{A}) &= \frac{e^{-H(\mathbf{W})}}{Z_{\mathbf{A}}} = \frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}_{\mathbf{A}}} e^{-H(\mathbf{W})} d\mathbf{W}} \\ &= \frac{e^{-\sum_{i<j} H_{ij}(w_{ij})}}{\int_{\mathbb{W}_{\mathbf{A}}} e^{-\sum_{i<j} H_{ij}(w_{ij})} d\mathbf{W}} \\ &= \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\left[\int_{m_{ij}}^{+\infty} e^{-H_{ij}(w_{ij})} dw_{ij}\right]^{a_{ij}}} = \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\zeta_{ij}^{a_{ij}}} \end{aligned} \tag{B3}$$

if $\mathbf{W} \in \mathbb{W}_{\mathbf{A}}$ and 0 otherwise—since each Hamiltonian considered in the present paper is separable, i.e., a sum of node pair-specific Hamiltonians: in formulas, $H(\mathbf{W}) = \sum_{i<j} H_{ij}(w_{ij})$.

APPENDIX C: CONDITIONAL NETWORK MODELS: ESTIMATING THE PARAMETERS

Let us now provide general expressions for the deterministic and the annealed recipe for parameter estimation. The first one follows from writing

$$\begin{aligned}\mathcal{L}_{\underline{\psi}} &= \ln Q(\mathbf{W}^*|\mathbf{A}^*) = -H(\mathbf{W}^*) - \ln Z_{\mathbf{A}^*}\\ &= -H(\mathbf{W}^*) - \ln\left[\int_{\mathbb{W}_{\mathbf{A}^*}} e^{-H(\mathbf{W})} d\mathbf{W}\right]\\ &= \sum_{i<j} H_{ij}(w_{ij}^*) - \ln\prod_{i<j}\zeta_{ij}^{a_{ij}} = \sum_{i<j}[H_{ij}(w_{ij}^*) - a_{ij}^*\ln\zeta_{ij}],\end{aligned} \tag{C1}$$

while the second one follows from writing

$$\begin{aligned}\mathcal{G}_{\underline{\psi}} &= \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\ln Q(\mathbf{W}^*|\mathbf{A}) = \langle\mathcal{L}_{\underline{\psi}}\rangle\\ &= \sum_{i<j}[H_{ij}(w_{ij}^*) - p_{ij}\ln\zeta_{ij}].\end{aligned} \tag{C2}$$

a. Scalar or homogeneous variant of the CEM. In the particular case of the UBRGM-induced, homogeneous variant of the CEM, one can derive the quenched distribution of the parameter β upon considering that it is a function of the discrete, random variable L. Since $L \sim \text{Bin}(N(N-1)/2, p)$, with $p = 2L^*/N(N-1)$, one finds that

$$\beta \sim \binom{\frac{N(N-1)}{2}}{W^*\beta} p^{W^*\beta}(1-p)^{\frac{N(N-1)}{2}-W^*\beta}, \tag{C3}$$

an expression allowing us to derive the expected value of β, i.e.,

$$\begin{aligned}\langle\beta\rangle &= \sum_{\beta=0}^{\frac{N(N-1)}{2W^*}} \beta\binom{\frac{N(N-1)}{2}}{W^*\beta} p^{W^*\beta}(1-p)^{\frac{N(N-1)}{2}-W^*\beta}\\ &= \frac{N(N-1)}{2W^*}p = \frac{\langle L\rangle}{W^*} = \frac{L^*}{W^*},\end{aligned} \tag{C4}$$

as well as its variance. Since

$$\begin{aligned}\langle\beta^2\rangle &= \sum_{\beta=0}^{\frac{N(N-1)}{2W^*}} \beta^2\binom{\frac{N(N-1)}{2}}{W^*\beta} p^{W^*\beta}(1-p)^{\frac{N(N-1)}{2}-W^*\beta}\\ &= \frac{N(N-1)}{2(W^*)^2}p + \frac{N(N-1)}{2(W^*)^2}\left[\frac{N(N-1)}{2(W^*)^2} - 1\right]p^2,\end{aligned} \tag{C5}$$

we have that

$$\begin{aligned}\text{Var}[\beta] &= \langle\beta^2\rangle - \langle\beta\rangle^2 = \frac{N(N-1)}{2(W^*)^2}p(1-p) = \frac{\text{Var}[L]}{(W^*)^2} =\\ &= \frac{L^*}{(W^*)^2}\left[\frac{N(N-1) - 2L^*}{N(N-1)}\right],\end{aligned} \tag{C6}$$

with $\text{Var}[L] = N(N-1)/2 \cdot p(1-p)$. Since the distribution obeyed by L converges to the normal distribution $\mathcal{N}(L^*, \text{Var}[L])$, the distribution obeyed by β converges to the distribution

$$\begin{aligned}g(\beta) &= \frac{W^*}{\sqrt{2\pi\,\text{Var}[L]}} e^{-\frac{(W^*\beta - L^*)^2}{2\text{Var}[L]}}\\ &= \frac{1}{\sqrt{2\pi\,\text{Var}[L]/(W^*)^2}} e^{-\frac{(\beta - L^*/W^*)^2}{2\text{Var}[L]/(W^*)^2}}\\ &= \frac{1}{\sqrt{2\pi\,\text{Var}[\beta]}} e^{-\frac{(\beta-\beta^*)^2}{2\text{Var}[\beta]}} = \mathcal{N}(\beta^*, \text{Var}[\beta]),\end{aligned} \tag{C7}$$

with $\beta^* = L^*/W^*$ and $\text{Var}[\beta] = \text{Var}[L]/(W^*)^2$.

In the case of the UBCM-induced homogeneous version of the CEM, L obeys the PB distribution, reading $\text{PB}(N(N-1)/2, \{p^{\text{UBCM}}\}_{i,j=1}^N)$ whose normal approximation reads $\mathcal{N}(L^*, \text{Var}[L])$, with $\text{Var}[L] = \sum_{i<j} p_{ij}^{\text{UBCM}}(1 - p_{ij}^{\text{UBCM}})$; as a consequence, the distribution obeyed by β converges to $\mathcal{N}(\beta^*, \text{Var}[\beta])$, with $\beta^* = L^*/W^*$ and $\text{Var}[\beta] = \text{Var}[L]/(W^*)^2$.

In the case of the LM-induced, homogeneous version of the CEM, L obeys the PB distribution reading $\text{PB}(N(N-1)/2, \{p^{\text{LM}}\}_{i,j=1}^N)$ whose normal approximation reads $\mathcal{N}(L^*, \text{Var}[L])$, with $\text{Var}[L] = \sum_{i<j} p_{ij}^{\text{LM}}(1 - p_{ij}^{\text{LM}})$; as a consequence, the distribution obeyed by β converges to $\mathcal{N}(\beta^*, \text{Var}[\beta])$, with $\beta^* = L^*/W^*$ and $\text{Var}[\beta] = \text{Var}[L]/(W^*)^2$.

b. Vector or weakly heterogeneous variant of the CEM. As pointed out in the main text, each annealed estimation overlaps with the average value of the related quenched distribution although (1) the latter ones are well separated, in the case of node 168, (2) only partly overlapped, in the case of node 171, and (3) the UBCM-induced and the LM-induced ones overlap while the UBRGM-induced one remains well separated, in the case of node 170 (see Fig. 4). Moreover, the deterministic estimation is always very close to the UBCM-induced annealed one—a result that may be a consequence of the accurate description of the empirical network topology provided by the UBCM—evidently, much more accurate than those provided by the UBRGM and the LM.

Each solid line in Fig. 4 represents a normal distribution whose average value and variance coincide with the ones of the corresponding sample distribution: although the empirical and theoretical CDFs seem to be in (very good) agreement, the Anderson-Darling test never rejects the normality hypothesis only for node 166 and does not reject the normality hypothesis in the case of the UBCM-induced distribution of values for node 168.

c. Tensor variant of the CEM. Let us now leave β_{ij} in its tensor form and constrain the set of weight-specific estimates $\hat{w}_{ij}, \forall\, i<j$. In this case, the three recipes lead to the following estimates:

$$\mathcal{L}_{\underline{\psi}} = \sum_{i<j}[-\beta_{ij}\hat{w}_{ij} + a_{ij}^*\ln\beta_{ij}] \Longrightarrow \beta_{ij} = \frac{a_{ij}^*}{\hat{w}_{ij}}, \tag{C8}$$

$$\mathcal{G}_{\underline{\psi}} = \sum_{i<j}[-\beta_{ij}\hat{w}_{ij} + p_{ij}\ln\beta_{ij}] \Longrightarrow \beta_{ij} = \frac{p_{ij}}{\hat{w}_{ij}}, \tag{C9}$$

$$\begin{aligned}\langle\beta_{ij}\rangle &= \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\beta_{ij}(\mathbf{A}) =\\ &= \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\frac{a_{ij}}{\hat{w}_{ij}} \quad\Longrightarrow\quad \langle\beta_{ij}\rangle = \frac{p_{ij}}{\hat{w}_{ij}},\end{aligned} \tag{C10}$$

a result signaling large differences between the deterministic recipe, on the one hand, and the quenched and annealed recipes, on the other—that, instead, coincide. If, however, $\hat{w}_{ij} \equiv w^*_{ij}$, $\forall\, i < j$ then, for consistency, $p_{ij} \equiv a^*_{ij}$ and the three recipes coincide.

d. Econometric'variant. As Figs. 3 and 5 show, the deterministic estimation is always quite different from the other two—the only exception being represented by the parameter α, under the UBCM-induced, binary recipe. Such a result should warn us from employing the deterministic estimation recipe *tout court*, as ignoring the variety of structures that are compatible with a given probability distribution $P(\mathbf{A})$ will, in general, affect the estimation of the parameters of interest.

[1] N. Vallarano, C. J. Tessone, and T. Squartini, Front. Phys. **8**, 286 (2020).

[2] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, and G. Caldarelli, Nat. Rev. Phys. **3**, 490 (2021).

[3] L. N. Ialongo, C. de Valk, E. Marchese, F. Jansen, H. Zmarrou, T. Squartini, and D. Garlaschelli, Sci. Rep. **12**, 11847 (2022).

[4] D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **93**, 188701 (2004).

[5] F. Schweitzer, G. Fagiolo, D. Sornette, and F. Vega-Redondo, Science **325**, 422 (2009).

[6] A. Fronczak and P. Fronczak, Phys. Rev. E **85**, 056113 (2012).

[7] P. R. Herman, Rev. World Econ. **158**, 127 (2022).

[8] J. Mullahy, J. Econom. **33**, 341 (1986).

[9] M. Burger, F. van Oort, and G. Linders, Spat. Econ. Anal. **4**, 167 (2009).

[10] J. Tinbergen, *Shaping the World Economy: Suggestions for an International Economic Policy* (The Twentieth Century Fund, New York, 1962).

[11] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).

[12] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, Hoboken, NJ, 1999).

[14] T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics* (Springer International Publishing, Cham, Switzerland, 2017).

[15] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, Nat. Rev. Phys. **1**, 58 (2019).

[16] M. Di Vece, D. Garlaschelli, and T. Squartini, Phys. Rev. Res. **4**, 033105 (2022).

[17] M. Di Vece, D. Garlaschelli, and T. Squartini, Chaos Solit. Fractals **166**, 112958 (2023).

[18] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).

[19] D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **102**, 038701 (2009).

[20] F. Parisi, T. Squartini, and D. Garlaschelli, New J. Phys. **22**, 053053 (2020).

[21] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).

[22] M. Conte, P. Cotterlaz, and T. Mayer, The CEPII Gravity Database, CEPII Working Paper 2022- 05 (2022).

[23] J.-H. Lin, E. Marchese, C. J. Tessone, and T. Squartini, Chaos Solit. Fractals **164**, 112620 (2022).

[24] T. Squartini, R. Mastrandrea, and D. Garlaschelli, New J. Phys. **17**, 023052 (2015).